ALSO: TR#12 Nonr-225(8[illegible]

M. V. JOHNS JR.

STANFORD UNIVERSITY

MULTIPLE TIME SERIES MODELLING

BY

EMANUEL PARZEN

TECHNICAL REPORT NO. 22

JULY 8, 1968

PREPARED UNDER GRANT DA-ARO(D)-31-124-G726

FOR

U. S. ARMY RESEARCH OFFICE

Department of Statistics

STANFORD, CALIFORNIA

MULTIPLE TIME SERIES MODELLING

By

Emanuel Parzen

TECHNICAL REPORT NO. 22

July 8, 1968

PREPARED UNDER GRANT DA-ARO(D)-31-124-G726

FOR

U. S. ARMY RESEARCH OFFICE

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

MULTIPLE TIME SERIES MODELLING

By

Emanuel Parzen
Professor of Statistics
Stanford, California

1. Introduction

Empirical multiple time series analysis is concerned with finding relations among r time series $X_1(\cdot), \ldots, X_r(\cdot)$, given finite samples

(1) $$\{X_1(t),\ t=1,2,\ldots,T\},\ \ldots,\ \{X_r(t),\ t=1,2,\ldots,T\} \quad .$$

Multiple time series modelling could be equivalently defined as multivariate analysis of a sample of dependent (rather than independent) random vectors

(2) $$\underline{X}(t) = \begin{bmatrix} X_1(t) \\ \vdots \\ X_r(t) \end{bmatrix} \quad .$$

We call $\underline{X}(\cdot) = \{X(t),\ t=0, \pm 1, \pm 2, \ldots\}$ a multiple time series.

The point of view that a multiple time series is a series of vectors (rather than a vector of series) seems useful for mathematical statistical investigations of the distribution of various sample statistics. Point One of this paper is: for pre-mathematical statistical investigations of the specification of the models to be fitted it may be essential to first model each component by itself.

This paper seeks to provide a general framework for the theory

---

Research supported by the Office of Naval Research under contract Nonr-225(80) and by the U. S. Army Research Office under grant DA-ARO(D)-31-124-G726. Reproduction in whole or in part permitted for any purpose of the United States Government. Presented at the Second International Symposium on Multivariate Analysis, Dayton, Ohio, on June 20, 1968.

and practice of multivariate analysis of time series. It seeks to compare:

(1) Spectral approaches to finding relations among time series.

(2) Time domain or innovations approaches to finding relations among time series.

The paper also seeks to focus attention on:

(3) Innovations approaches to cross-spectral estimation.

(4) The problem of multivariate analysis of the joint innovations covariance matrix and the sampling properties of its estimators.

The various sections are entitled: 2. Innovation Approaches to Modelling, 3. Spectral Approaches to Modelling, 4. Relations Between Time Series, 5. Autoregressive Approach to a Single Series, 6. Multiple Spectral Density Estimation.

2. Innovation Approaches to Modelling

When we admit the possibility that our vector samples $\underline{X}(1), \ldots, \underline{X}(T)$ are not independent, and seek to build stochastic dynamic models, the statistical inference problem could be conceived as one of estimating

$$(1) \qquad m_j(t) = E[X_j(t)] \quad ,$$

$$(2) \qquad K_{jk}(s,t) = \mathrm{Cov}[X_j(s), X_k(t)] \quad ;$$

these means and covariances specify the probability law of the observations when it is assumed to be multivariate normal. In order that there not be as many or more parameters as observations, one must assume models which restrict $m_j(\cdot)$ and $K_{jk}(\cdot,\cdot)$, since otherwise statistical inference is impossible.

A multiple time series $\underline{X}(\cdot)$ is called <u>covariance stationary</u> [see Parzen (1962), Chapter 3] if for each index j and k there is a function $R_{jk}(v)$ of $v = 0, \pm 1, \ldots$ such that

$$(3) \qquad \mathrm{Cov}[X_j(s), X_k(t)] = R_{jk}(t-s) \quad .$$

The r by r matrix

$$(4) \quad \underline{R}(v) = \begin{bmatrix} R_{11}(v) & \ldots & R_{1r}(v) \\ \ldots & \ldots & \ldots \\ R_{r1}(v) & \ldots & R_{rr}(v) \end{bmatrix}, \qquad R_{hj}(v) = \mathrm{Cov}[X_h(t), X_j(t+v)] \quad ,$$

is called the <u>covariance matrix</u> $\underline{R}(\cdot)$ of the <u>covariance stationary multiple</u> time series $\{\underline{X}(t), t = 1, 2, \ldots, T\}$.

The sample statistics appropriate for inferring models for

covariance stationary time series are often interpretable even for non-stationary time series [either as time varying statistics, as in Priestley (1965), or through transformations to stationarity, as in Parzen (1967a) or Whittle (1963b)]. Therefore we assume that the multiple time series $\underline{X}(\cdot)$ being discussed is covariance stationary.

Let us review briefly models for a univariate stationary time series $X(\cdot)$; its covariance function $R(v)$ has spectral representation

(5)
$$R(v) = \int_{-\pi}^{\pi} \cos v\omega \; dF(\omega) \quad .$$

When seeking to model a time series $X(\cdot)$ with given covariance function $R(\cdot)$ and spectral distribution function $F(\cdot)$, in principle one may treat separately the three types of distribution functions into which $F(\cdot)$ may be decomposed:

(6)
$$F(\omega) = F_d(\omega) + F_{ac}(\omega) + F_{sc}(\omega) \quad ;$$

in words, $F(\cdot)$ is the sum of three distribution functions which are, respectively, discrete (or purely discontinuous), absolutely continuous, and singular continuous.

Observed time series are assumed to have a <u>mixed spectrum</u>, in the sense that: (i) the singular continuous part of the spectral distribution function vanishes, (ii) the discrete part has only a finite number of jumps, and (iii) the absolutely continuous part has a spectral density function $f(\omega)$ satisfying

(7)
$$\int_{-\pi}^{\pi} \log f(\omega) d\omega > -\infty \quad .$$

Note that $f(\omega)$ is an even non-negative function such that

(8) $$F_{ac}(\omega) = \int_{-\pi}^{\omega} f(\omega')d\omega' .$$

We call $m(t) = E[X(t)]$ the mean value function of $X(\cdot)$. It may be shown that $X(t) - m(t)$ may be written as the sum of two time series,

(9) $$X(t) - m(t) = X_d(t) + X_c(t)$$

satisfying

(10) $$X_d(t) = A_0 + \sum_{j=1}^{r} \{A_j \cos \lambda_j t + B_j \sin \lambda_j t\}$$

where A_0, A_j, B_j are uncorrelated random variables and λ_j are frequencies in the band $0 < \lambda_j \leq \pi$, while

(11) $$X_c(t) = \eta(t) + b_1\, \eta(t-1) + b_2\, \eta(t-2) + \ldots$$

where $\{b_\nu\}$ are constants such that $\Sigma\, b_\nu^2 < \infty$ and $\{\eta(\nu)\}$ are uncorrelated random variables.

The probability distribution of A_0, A_j, B_j cannot be estimated from a single realization of $X(\cdot)$, or even of $X_d(\cdot)$; all one can hope to estimate is the value of these variables in the realization observed. Thus A_0, A_j, B_j can be treated as constants (rather than random variables) for purposes of statistical inference and $X_d(\cdot)$ can be treated as part of the mean value function of $X(\cdot)$. The mean value function of $X(\cdot)$ has to be eliminated by some detrending procedure (which could involve spectral analysis) in order to do statistical inference on $X_c(\cdot)$, the "fluctuation" part of $X(\cdot)$.

Point Two of this paper is: in multiple time series modelling we

can assume that we are dealing with zero mean jointly covariance stationary time series $X_j(\cdot)$, each satisfying (7) and therefore satisfying a model of the form

$$X_j(t) = \eta_j(t) + b_1^{(j)} \eta_j(t-1) + b_2^{(j)} \eta_j(t-2) + \ldots \quad . \tag{12}$$

To understand the meaning of the random variables $\eta_j(t)$, let us hereafter consider <u>normal</u> time series $\{X(t)\}$. One can associate to a univariate series $X(\cdot)$ a series of successive conditional expectations (or minimum mean square error predictors)

$$X^*(t) = E[X(t)|X(t-1), X(t-2), \ldots] \quad , \tag{13}$$

and conditional variances (or mean square prediction errors)

$$\tilde{\sigma}_t^2 = \mathrm{Var}[X(t)|X(t-1), X(t-2), \ldots] = E[|X(t) - X^*(t)|^2] \quad . \tag{14}$$

For non-normal time series, the notion of projection is used in place of conditional expectation; see Rozanov (1967).

The one step prediction error, denoted

$$\eta(t) = X(t) - X^*(t) \quad \text{or} \quad \tilde{X}(t) = X(t) - X^*(t) \quad , \tag{15}$$

is called the <u>innovation</u> at time t. The successive innovations $\eta(t)$ are a sequence of uncorrelated (independent when $X(\cdot)$ is normal, as assumed here) random variables,

$$E[\eta(s)\ \eta(t)] = 0 \qquad \text{if} \quad s \neq t \quad . \tag{16}$$

An uncorrelated sequence $\eta(\cdot)$ is called <u>white noise</u>; if all variances are equal it is called stationary white noise.

Writing $X(t)$ as an infinite series in $\eta(t)$, $\eta(t-1)$, ... is one way of expressing the time series $X(\cdot)$ as the output of a filter whose input is white noise $\eta(\cdot)$:

$$\eta(\cdot) \longrightarrow \boxed{\text{Filter } \Phi} \longrightarrow X(\cdot) = \Phi[X(\cdot)] \ .$$

By representing a time series as the output of a filter whose input is <u>innovation</u> white noise we are able to conveniently solve estimation (prediction, signal extraction) problems and simulation problems for the time series.

For a univariate time series $X(\cdot)$ which is assumed to be normal, covariance stationary, have zero means, and non-deterministic [in the sense that it satisfies model (11)], the modelling problem can be solved by estimating either:

(i) its covariance function $R(v)$, or

(ii) its spectral density function $f(\omega)$, or

(iii) its innovation variance $\tilde{\sigma}^2$ and the filter $\Phi[\cdot]$ which transforms $\eta(\cdot)$ to $X(\cdot)$.

Point Three of this paper is: approach (iii) is the most satisfactory for two reasons: (1) as the answer we seek since it is the most convenient form for prediction and control, (2) as the most suitable means of obtaining (ii) [the details of how to do this are discussed in Section 5].

Point Four of this paper is: for a multiple stationary normal time series two types of innovation approaches to modelling can be considered:

(i) the individual innovation approach, and

(ii) the joint innovation approach.

The individual innovations, denoted $\eta_j(t)$, are defined in terms of the predictors of each series given its own past

(17) $$X_j^*(t) = E[X_j(t) | X_j(s),\ s < t]$$

by

(18) $$\eta_j(t) = X_j(t) - X_j^*(t) \quad .$$

Equation (12) is a representation of $X_j(t)$ in terms of its individual innovations.

The joint innovations, denoted $\underline{\eta}_j(t)$, are defined in terms of the predictors, denoted $\underline{X}_j^*(t)$, of each series given the pasts of all of them; in symbols

(19) $$\underline{X}_j^*(t) = E[X_j(t) | X_k(s),\ s < t \ \text{ and } \ k = 1, 2, \ldots, r]$$

and

(20) $$\underline{\eta}_j(t) = X_j(t) - \underline{X}_j^*(t) \quad .$$

The joint innovation multiple time series, denoted $\underline{\eta}(\cdot)$, and defined by $\underline{\underline{\eta}}(\cdot)' = (\underline{\eta}_1(\cdot), \ldots, \underline{\eta}_r(\cdot))$, is multiple white noise in the sense that

(21) $$E[\underline{\underline{\eta}}(s)\ \underline{\underline{\eta}}'(t)] = 0 \qquad \text{for } s \neq t \quad ,$$

and therefore is described by the innovation covariance matrix

(22) $$\tilde{\Sigma} = E[\underline{\underline{\eta}}(t)\ \underline{\underline{\eta}}'(t)] \quad .$$

The joint innovation approach models $\underline{X}(\cdot)$ by estimating: (i) the innovation covariance matrix $\tilde{\Sigma}$, and (ii) the multi input- multi output

filter which transforms the joint innovations $\underline{\underline{\eta}}(\cdot)$ to the observed multiple time series $\underline{X}(\cdot)$.

The individual innovation approach models $\underline{X}(\cdot)$ by estimating: (i) each individual innovation series $\eta_j(\cdot)$ and the filter transforming $\eta_j(\cdot)$ to $X_j(\cdot)$, (ii) the multiple time series [denoted $\underline{\eta}_{ind}(t)$ and defined by $\underline{\eta}_{ind}(\cdot)' = (\eta_1(\cdot), \ldots, \eta_r(\cdot))$] of individual innovations in terms of their joint innovations [to be denoted $\underline{\epsilon}(\cdot)$ and called the <u>innovation innovations</u>] and the multi input - multi output filter which transforms $\underline{\epsilon}(\cdot)$ to $\underline{\eta}_{ind}(\cdot)$.

Point Five of this paper is: to estimate the joint innovation structure of a multiple time series we fit it by a <u>sufficiently long</u> joint autoregressive scheme. Probabilistic justification of such fits is provided by the work of Masani (see, for example, Section 13 of Masani's review paper (1966) at the First Symposium on Multivariate Analysis).

A zero mean covariance stationary multiple time series $\underline{X}(\cdot)$ is called a <u>joint autoregressive scheme of order</u> m if the <u>infinite memory</u> predictor $\underline{X}^*(t)$ can be expressed as a linear combination of $\underline{X}(t-1), \ldots, \underline{X}(t-m)$:

$$\underline{X}^*(t) = A(1)\,\underline{X}(t-1) + \ldots + A(m)\,\underline{X}(t-m) \quad . \tag{23}$$

When the multiple time series $\underline{X}(\cdot)$ is known to be a joint autoregressive scheme of order m, the autoregressive matrix coefficients $A(1), \ldots, A(m)$ are estimated from a sample $\{\underline{X}(t),\ t = 1, 2, \ldots, T\}$ of size T by the solutions $\hat{A}(1), \ldots, \hat{A}(m)$ of a system of equations called the multiple Yule-Walker equations:

$$(24) \qquad \sum_{j=1}^{m} \hat{A}(j)\, \underline{R}_T(j-k) = \underline{R}_T(-k)\,, \qquad k = 1, 2, \ldots, m\,,$$

where $\underline{R}_T(v)$ is the sample covariance matrix defined in the next section.

The Yule-Walker equations are suggested by the fact that the matrix coefficients $A(1), \ldots, A(m)$ of the finite memory predictor in (23) satisfy the <u>normal</u> equations

$$(25) \quad E[\underline{X}^*(t)\, \underline{X}'(t-k)] = E[\underline{X}(t)\, \underline{X}'(t-k)]\,, \qquad k = 1, 2, \ldots, m$$

or

$$(26) \qquad \sum_{j=1}^{m} A(j)\, \underline{R}(j-k) = \underline{R}(-k)\,, \qquad k = 1, 2, \ldots, m\,.$$

The prediction error covariance matrix $\tilde{\Sigma}$ satisfies

$$(27) \qquad \tilde{\Sigma} = [\{\underline{X}(t) - \underline{X}^*(t)\}\, \underline{X}'(t)] = \underline{R}(0) - \sum_{j=1}^{m} A(j)\, \underline{R}(j)\;;$$

the natural estimator of $\tilde{\Sigma}$ is

$$(28) \qquad \hat{\tilde{\Sigma}} = \underline{R}_T(0) - \sum_{j=1}^{m} \hat{A}(j)\, \underline{R}_T(j)\,.$$

It is important to note that the <u>computation</u> of $\hat{A}(1), \ldots, \hat{A}(m), \hat{\tilde{\Sigma}}$ is most conveniently done <u>recursively</u> [as in Whittle (1963), Jones (1964), or Robinson (1967)].

From the work of Wold (1954), Mann and Wald (1944) and Whittle (1953) we know the properties of the autoregressive coefficient estimators $\hat{A}(1), \ldots, \hat{A}(m)$; indeed, their properties are very similar to those of estimators of multivariate regression coefficients [as given, for example, in Kendall and Stuart (1966), p. 275].

What has not been explicitly proved in the literature, but seems plausible (and basic to the innovation approach to modelling) is that $(T\text{-}m)\ \hat{\tilde{\Sigma}}$ is approximately distributed as the Wishart distribution of dimension r, degrees of freedom $T - m$, and covariance matrix $\tilde{\Sigma}$. In Section 3, we will make the point that multivariate analysis of $\hat{\tilde{\Sigma}}$ is an important tool of multiple time series modelling.

3. Spectral Approaches to Modelling

The spectral approach to multiple stationary time series analysis assumes that each component is non-deterministic, and in addition assumes that the covariance matrix $\underline{R}(\cdot)$ is absolutely summable in the sense that

$$(1) \qquad \sum_{v} |R_{hj}(v)| < \infty \qquad \text{for } h, j = 1, 2, \ldots, r \quad ,$$

so that an explicit formula can be given for the spectral density matrix

$$(2) \qquad \underline{f}(\omega) = \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} e^{-iv\omega} \underline{R}(v)$$

or

$$(3) \quad \underline{f}(\omega) = \begin{bmatrix} f_{11}(\omega) & \ldots & f_{1r}(\omega) \\ \ldots & \ldots & \ldots \\ f_{r1}(\omega) & \ldots & f_{rr}(\omega) \end{bmatrix}, \qquad f_{hj}(\omega) = \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} e^{-iv\omega} R_{hj}(v) \quad ;$$

in terms of $\underline{f}(\omega)$, we have a Fourier representation for $\underline{R}(v)$:

$$(4) \qquad \underline{R}(v) = \int_{-\pi}^{\pi} e^{iv\omega} \underline{f}(\omega) d\omega \quad .$$

One calls $\underline{f}(\omega)$ the <u>spectral density matrix</u> of the covariance stationary multiple time series $\underline{X}(\cdot)$; for further discussion see Rozanov (1967), Granger (1964), Jenkins and Watts (1968).

Point Six of this paper is: there are <u>three kinds</u> of sample statistics in multiple time series modelling, which one should use simultaneously and between which one should know how to transform quickly. The three kinds of sample statistics are:

(i) the sample covariance matrix,

(ii) a matrix of estimated spectral densities,

(iii) various innovations and sample autoregressive coefficients.

The sample covariance matrix is defined by

$$(5) \qquad \underline{R}_T(v) = \frac{1}{T} \sum_{t=1}^{T-v} \underline{X}(t)\, \underline{X}'(t+v) , \qquad \text{for} \quad v = 0, 1, \ldots, T-1 .$$

When the multiple time series $\underline{X}(\cdot)$ has zero means and is covariance stationary, one regards $\underline{R}_T(v)$ as an estimate of the value of the covariance matrix $\underline{R}(v)$. Since $\underline{R}(-v) = \underline{R}'(v)$ we define $\underline{R}_T(-v) = \underline{R}_T'(v)$.

It should be noted that in the course of our discussion of estimated spectral densities, it will be seen that in practice one should <u>rarely</u> use (5) to <u>compute</u> $\underline{R}_T(v)$, and one should <u>not usually</u> use (5) to even <u>define</u> $\underline{R}_T(v)$ for "detrended" time series.

While the sample covariance matrix $\underline{R}_T(v)$ is of interest to determine the lags v at which the components of $\underline{R}_T(v)$ are significantly non-zero, <u>for time series modelling</u> $\underline{R}_T(\cdot)$ <u>needs to be transformed either to spectral density estimates or autoregressive coefficient estimates.</u>

A matrix $\hat{\underline{f}}(\omega)$ of estimated spectral densities $\hat{f}_{hj}(\omega)$ is denoted

$$(6) \qquad \hat{\underline{f}}(\omega) = \begin{bmatrix} \hat{f}_{11}(\omega) & \hat{f}_{12}(\omega) & \ldots & \hat{f}_{1r}(\omega) \\ \hat{f}_{21}(\omega) & \hat{f}_{22}(\omega) & \ldots & \hat{f}_{2r}(\omega) \\ \ldots & \ldots & \ldots & \ldots \\ \hat{f}_{r1}(\omega) & \hat{f}_{r2}(\omega) & \ldots & \hat{f}_{rr}(\omega) \end{bmatrix} .$$

Point Seven of this paper is: the spectral approach to multivariate analysis of time series may be defined to be concerned with the relations among time series that can be inferred statistically from the matrix of

estimated spectral densities [as well as probabilistically from the matrix of true spectral densities; for example, see Koopmans (1964)].

We briefly describe ways of forming estimated spectra $\hat{\underline{f}}(\omega)$ [for a tabular presentation of these remarks see Parzen (1968)]. First, form the sample Fourier transform

$$(7) \qquad \underline{Z}(\omega) = \sum_{t=1}^{T} e^{-i\omega t}\, \underline{X}(t)$$

at specified frequencies ω (we prefer $0, \frac{2\pi}{2T}, \ldots, (2T-1)\frac{2\pi}{2T}$). Second, at these frequencies form the sample spectral density matrix by the formula

$$(8) \qquad \underline{f}_T(\omega) = \frac{1}{2\pi T}\, \underline{Z}(\omega)\, Z'(-\omega)$$

which satisfies the relations

$$(9) \qquad \underline{R}_T(v) = \int_{-\pi}^{\pi} e^{iv\omega}\, \underline{f}_T(\omega)\, d\omega \quad ,$$

$$(10) \qquad \underline{f}_T(\omega) = \frac{1}{2\pi} \sum_{|v| < T} e^{-iv\omega}\, \underline{R}_T(v) \quad .$$

Third, form estimators of $\underline{f}(\omega)$ by averaging adjacent values of $\underline{f}_T(\omega)$; this averaging process is computationally faster if the averages one considers are of the form

$$(11) \qquad \hat{\underline{f}}(k\,\tfrac{2\pi}{T}) = \Sigma\, K_T((j-k)\,\tfrac{2\pi}{T})\, \underline{f}_T(j\,\tfrac{2\pi}{T}) \quad ,$$

which we call filtered sample spectral density functions.

An alternative third step, which seems to me the most convenient (and because of Fast Fourier Transform techniques perhaps faster) way to

compute $\underline{\hat{f}}(\cdot)$, is via the method of covariance averages [compare Parzen (1967b) or Jenkins (1967)]

$$(12) \qquad \underline{\hat{f}}(\omega) = \frac{1}{2\pi} \sum_{|v|<T} e^{-iv\omega} k\left(\frac{v}{M}\right) \underline{R}_T(v)$$

in terms of a suitable <u>kernel</u> $k(\cdot)$ and constant M called the <u>truncation point</u>. We prefer (12) to (11) since one can readily compute $\underline{R}_T(v)$ for $v = 0, 1, \ldots, T - 1$ (through the Fast Fourier Transform) by the formula

$$(13) \qquad \underline{R}_T(v) = \frac{2\pi}{2T} \sum_{k=0}^{2T-1} \exp\left(ivk \frac{2\pi}{2T}\right) \underline{f}_T\left(k \frac{2\pi}{2T}\right) .$$

For the proof of (13), compare Gentleman and Sande (1966), p. 573.

To interpret estimated spectra, one has to take account of both their variability and their bias. The basic approximation on variability [which was first noted by Goodman (1963) and proved by Wahba (1968) and Brillinger (1970)] is that an estimated spectral density matrix $\underline{\hat{f}}(\omega)$ of form (12), or equivalently (11), has the following approximate distribution: $\nu\, \underline{\hat{f}}(\omega)$ has a complex Wishart distribution of dimension r, degrees of freedom ν, and covariance matrix $\underline{f}(\omega)$, where

$$(14) \qquad \frac{1}{\nu} = \frac{M}{T} \int_{-\infty}^{\infty} k^2(u)\,du .$$

By identifying the distribution of $\underline{\hat{f}}(\omega)$ with the Wishart distribution, one reduces to <u>standard problems of multivariate analysis</u> the problem of finding the distribution of various statistics derived from $\underline{\hat{f}}(\omega)$.

To conclude this section, we note that the foregoing computational path seems especially appropriate when one cannot assume the observed

time series to have zero mean (no trend), and desires to detrend the series $\underline{X}(\cdot)$ by passing each component $X_j(\cdot)$ through a filter to form a detrended series ${}_dX_j(\cdot)$:

(15) $$ {}_dX_j(t) = \sum_{\alpha=-m}^{n} X_j(t-\alpha)\, w_j(\alpha) \quad . $$

However the Fourier transform ${}_d\underline{Z}_j(\cdot)$ can be formed [without first forming ${}_d\underline{X}(\cdot)$] directly from the Fourier transform $\underline{Z}(\cdot)$ by

(16) $$ {}_dZ_j(\omega) = Z_j(\omega)\, W_j(\omega) $$

where

(17) $$ W_j(\omega) = \sum_{\alpha} w_j(\alpha)\, \exp(i\omega\alpha) \quad . $$

Further to form ${}_d\underline{X}(\cdot)$ as a function of time one need only invert the Fourier transform of ${}_d\underline{X}(\cdot)$. The sample spectral density matrix and sample covariance matrix of a detrended multiple time series ${}_d\underline{X}(\cdot)$ seem to me to be best computed by

(18) $$ \underline{f}_T(\omega) = \frac{1}{2\pi T}\, {}_d\underline{Z}(\omega)\, {}_d\underline{Z}'(-\omega) $$

and (13) respectively. I must admit that as yet I have no practical experience with comparisons of formula (18) with more "direct" methods of calculation.

4. Relations Between Time Series

Given two jointly stationary zero mean multiple time series $\underline{X}(\cdot)$ and $\underline{Y}(\cdot)$ there are a variety of relation filters. To regard $\underline{X}(\cdot)$ as the independent variable is to regard it as the input of a filter whose output, denoted $\underline{Y}^*(\cdot)$, provides a representation of $\underline{Y}(\cdot)$ as a sum of two terms:

$$(1) \qquad \underline{Y}(t) = \underline{Y}^*(t) + \underline{\epsilon}(t) , \quad \underline{\epsilon}(t) = \underline{Y}(t) - \underline{Y}^*(t) .$$

Point Eight of this paper is: it is most meaningful to define $\underline{Y}^*(t)$ as a minimum mean square error linear predictor of $\underline{Y}(t)$ given a specified past of the $\underline{X}(\cdot)$ series [for example $\{\underline{X}(s),\ s = 0, \pm 1, \pm 2, \ldots\}$ or $\{\underline{X}(s),\ s \le t\}$]. In other words, $E[|\underline{Y}(t) - \underline{Y}^*(t)|^2]$ is a minimum among all possible linear functionals of the specified values of $X(\cdot)$. Then $\underline{\epsilon}(\cdot)$ is the error series, characterized by the normal property (for each $t = 0, \pm 1, \pm 2, \ldots$)

$$(2) \qquad E[\underline{X}(s)\ \underline{\epsilon}'(t)] = 0 \quad \text{for all indices } s \text{ such that } \underline{X}(s) \text{ is part of the memory used to form } \underline{Y}^*(t).$$

In addition to specifying the past of $\underline{X}(\cdot)$ used to form $\underline{Y}^*(t)$ as a linear predictor of $\underline{Y}(t)$, one may specify "matrix restraints" on the form of $\underline{Y}^*(t)$ of the type considered by Brillinger (1969) in his paper at this symposium.

The system which transforms $\underline{X}(\cdot)$ to $\underline{Y}^*(\cdot)$ is a filter which when $\underline{X}(\cdot)$ and $\underline{Y}(\cdot)$ are jointly covariance stationary is time invariant. The spectral representation of this filter is a matrix function of ω, denoted $\underline{B}_{Y^*}(\omega)$ and called the filter transfer function,

best described by assuming that the filter is an infinite moving average

$$(3) \qquad \underline{Y}^*(t) = \sum_{k=-\infty}^{\infty} \underline{\beta}(k)\ \underline{X}(t-k)$$

where $\{\underline{\beta}(k),\ k = 0, \pm 1, \ldots\}$ is a sequence of matrices called the <u>filter response</u> function. The filter transfer function is defined by

$$(4) \qquad \underline{B}_{Y^*}(\omega) = \sum_{k=-\infty}^{\infty} e^{-i\omega k}\ \underline{\beta}(k) \quad .$$

The relation between $\underline{Y}(\cdot)$ and $\underline{X}(\cdot)$ is resolved into a <u>deterministic dynamic system</u> represented by the relation filter with filter transfer function $\underline{B}_{Y^*}(\cdot)$ and a <u>stochastic driving function</u> represented by the error series $\underline{\epsilon}(\cdot)$ with spectral density matrix denoted by $\tilde{f}_{Y^*}(\omega)$.

When $\underline{Y}^*(t)$ is a function of all values of $\underline{X}(\cdot)$ in the sense that

$$(5) \qquad \underline{Y}^*(t) = E[\underline{Y}(t) \,|\, \underline{X}(s),\ s = 0, \pm 1, \ldots] \ ,$$

we call $\tilde{\underline{f}}_{Y^*}(\omega)$ the <u>partial spectral density</u> matrix of $\underline{Y}(\cdot)$, given $\underline{X}(\cdot)$.

Point Nine of this paper is: the spectral theory of multivariate analysis of time series has been mainly concerned with finding:
(i) formulas for $\underline{B}_{Y^*}(\omega)$ and $\tilde{\underline{f}}_{Y^*}(\omega)$ in terms of the joint spectral density matrix of the $\underline{X}(\cdot)$ and $\underline{Y}(\cdot)$ multiple time series

$$(6) \qquad \underline{f}(\omega) = \begin{bmatrix} \underline{f}_{XX}(\omega) & \underline{f}_{XY}(\omega) \\ \underline{f}_{YX}(\omega) & \underline{f}_{YY}(\omega) \end{bmatrix} ,$$

and (ii) the sampling properties of the natural estimators $\hat{\underline{B}}_{Y^*}(\omega)$ and $\hat{\tilde{\underline{f}}}_{Y^*}(\omega)$ formed from an estimated spectral density matrix

$$(7) \qquad \hat{\underline{f}}(\omega) = \begin{bmatrix} \hat{\underline{f}}_{XX}(\omega) & \hat{\underline{f}}_{XY}(\omega) \\ \hat{\underline{f}}_{YX}(\omega) & \hat{\underline{f}}_{YY}(\omega) \end{bmatrix} = \begin{bmatrix} \hat{\underline{f}}_{XX}(\omega) & \hat{\underline{f}}_{XY}(\omega) \\ \hat{\underline{f}}'_{XY}(\omega) & \hat{\underline{f}}_{YY}(\omega) \end{bmatrix}$$

under the assumption that for a suitable number ν of degrees of freedom $\nu\, \hat{\underline{f}}(\omega)$ obeys a complex Wishart distribution of ν degrees of freedom and covariance matrix $\underline{f}(\omega)$.

For example, by the usual matrix pivoting procedures used to solve normal equations, one can transform [see Parzen (1967b)] an estimated spectral density matrix (6) to a partitioned matrix

$$(8) \qquad \begin{bmatrix} \hat{\underline{f}}^{-1}_{XX}(\omega) & \hat{\underline{B}}'_{Y^*}(\omega) \\ -\bar{\hat{\underline{B}}}_{Y^*}(\omega) & \hat{\tilde{\underline{f}}}_{Y^*}(\omega) \end{bmatrix}$$

where

$$(9) \qquad \begin{aligned} \bar{\hat{\underline{B}}}_{Y^*}(\omega) &= \hat{\underline{f}}_{YX}(\omega)\, \hat{f}^{-1}_{XX}(\omega) \\ \hat{\tilde{\underline{f}}}_{Y^*}(\omega) &= \hat{\underline{f}}_{YY}(\omega) - \bar{\hat{\underline{B}}}_{Y^*}(\omega)\, \hat{\underline{f}}_{XY}(\omega) \\ &= \hat{\underline{f}}_{YY}(\omega) - \hat{\underline{f}}_{YX}(\omega)\, \underline{f}^{-1}_{XX}(\omega)\, \bar{\hat{\underline{f}}}'_{YX}(\omega) \end{aligned}$$

are natural estimators of the regression transfer function and error spectrum respectively for $\underline{Y}^*(\cdot)$ defined by (5); for the multivariate analogue of (9) see Anderson (1958).

The work that has been done on estimating relations between time series in terms of $\underline{B}_{Y^*}(\omega)$ and $\underline{f}_{Y^*}(\omega)$ leaves open a number of problems and issues which it is the <u>major aim</u> of this paper to point out:

(1) One would like to describe in the time domain the filter which $\hat{\underline{B}}_{Y^*}(\omega)$ estimates in the frequency domain; one way of doing this is to write

$$(10) \qquad \hat{\underline{B}}_{Y^*}(\omega) = \sum_{k=-m}^{n} \underline{\beta}(k) e^{-i\omega k} + \{\hat{\underline{B}}_{Y^*}(\omega) - \underline{B}_Y(\omega)\}$$

where m, n, and $\underline{\beta}(k)$ are to be estimated and the "errors" $\hat{\underline{B}}_{Y^*}(\omega) - \underline{B}_Y(\omega)$ are approximately normal with zero means and asymptotic variances that can be estimated; often the "errors" at different frequencies can be shown to be asymptotically independent. Pioneering and elegant work on this problem has been done by E. J. Hannan [see his papers, Hannan (1963), (1965), (1967), Hamon and Hannan (1963)]. From (10) one estimates the coefficients $\underline{\beta}(k)$ by regression analysis.

(2) One would like to describe (model) in terms of a time domain filter with white noise input the error series $\underline{\epsilon}(\cdot)$ whose spectral density matrix $\tilde{\underline{f}}_{Y^*}(\omega)$ is estimated by $\hat{\tilde{\underline{f}}}_{Y^*}(\omega)$.

(3) The sampling theory of the usual spectral estimators (namely, smoothed sample spectral densities) is based entirely on variability theory [for example, Rosenblatt (1959) and Goodman (1963)] and ignores the fact that estimation of cross-spectra by the usual method of smoothed sample spectral densities is subject to serious bias errors [Akaike and Yamanouchi (1961), Nettheim (1966), Parzen (1967b), Tick (1967)]. I believe that it can be shown that spectral estimators which have "minimum" bias and variability can be found by fitting long enough autoregressive schemes. We indicate in this paper several "autoregression approaches" to cross-spectral estimation and to fitting time domain models to time series.

(4) The relations between time series which can be inferred from estimated spectra are not "causal" unless the relations are between time series physically measured at the input and output respectively of a causal filter. Causal relations can be fitted only through "innovations" which can be found by fitting long autoregressive schemes. In other words, for finding relations between two arbitrary time series, spectral methods suffer from the drawback that they work directly only for predictors whose memory involves the future as well as the past. They cannot easily be used to estimate the error spectrum, and (more importantly) the filter transfer function from $\underline{X}(\cdot)$ to $\underline{Y}^*(\cdot)$, for cases such as the following:

$$
\begin{aligned}
\underline{Y}^*(t) &= E[\underline{Y}(t) \,|\, \underline{X}(s),\ s < t] , \\
\underline{Y}^*(t) &= E[\underline{Y}(t) \,|\, \underline{X}(s),\ s < t \ \text{ and } \ \underline{Y}(s),\ s < t] , \qquad (11) \\
\underline{Y}^*(t) &= E[\underline{Y}(t) \,|\, \underline{X}(s),\ s \leq t \ \text{ and } \ \underline{Y}(s),\ s < t] .
\end{aligned}
$$

Autoregressive methods seem to provide directly estimators of these semi-infinite memory predictors.

To the multiple time series

$$
\begin{bmatrix} \underline{X}(t) \\ \underline{Y}(t) \end{bmatrix} \qquad (12)
$$

one can fit a sufficiently long autoregressive scheme

$$
\begin{bmatrix} \underline{X}(t) \\ \underline{Y}(t) \end{bmatrix} = A(1) \begin{bmatrix} \underline{X}(t-1) \\ \underline{Y}(t-1) \end{bmatrix} + \ldots + A(m) \begin{bmatrix} \underline{X}(t-m) \\ \underline{Y}(t-m) \end{bmatrix} + \underline{\underline{\eta}}(t) \qquad (13)
$$

where $\underline{\underline{\eta}}(t)$ is multiple white noise. Writing

$$(14) \qquad A(j) = \begin{bmatrix} A_{XX}(j) & A_{XY}(j) \\ A_{YX}(j) & A_{YY}(j) \end{bmatrix}, \qquad \underline{\underline{\eta}}(t) = \begin{bmatrix} \eta_X(t) \\ \eta_Y(t) \end{bmatrix}$$

one obtains a relation between the $\underline{X}(\cdot)$ and $\underline{Y}(\cdot)$ series:

$$(15) \qquad \begin{aligned} &\underline{Y}(t) - A_{YY}(1)\,\underline{Y}(t-1) - \dots - A_{YY}(m)\,\underline{Y}(t-m) \\ &= A_{YX}(1)\,\underline{X}(t-1) - \dots - A_{YX}(m)\,\underline{X}(t-m) + \eta_Y(t) \quad . \end{aligned}$$

We next show how to add $\underline{X}(t)$ to the relation (15). Given $\underline{X}(t)$, and the past of $X(\cdot)$ and $Y(\cdot)$ up to $t - 1$, one can form

$$(16) \qquad \begin{aligned} \eta_X(t) &= \underline{X}(t) - \underline{X}^*(t) \\ &= \underline{X}(t) - A_{XX}(1)\,\underline{X}(t-1) - \dots - A_{XX}(m)\,\underline{X}(t-m) \\ &\quad - A_{XY}(1)\,\underline{Y}(t-1) - \dots - A_{XY}(m)\,\underline{Y}(t-m) \quad . \end{aligned}$$

Next, from $\eta_X(t)$ and the innovation covariance matrix $\tilde{\Sigma}$ one can form a predictor $\eta_Y^*(t)$ of $\eta_Y(t)$. To write $\eta_Y^*(t)$ explicitly, partition $\tilde{\Sigma}$:

$$(17) \qquad \tilde{\Sigma} = \begin{bmatrix} \tilde{\Sigma}_{XX} & \tilde{\Sigma}_{XY} \\ \tilde{\Sigma}_{YX} & \tilde{\Sigma}_{YY} \end{bmatrix} \quad .$$

Then

$$(18) \qquad \eta_Y^*(t) = \tilde{\Sigma}_{YX}\,\tilde{\Sigma}_{XX}^{-1}\,\eta_X(t) \quad .$$

In (15) substitute $\eta_Y^*(t)$, given by (18), for $\eta_Y(t)$; one thus obtains the formula

$$E[\underline{Y}(t)|\underline{X}(s),\ s \le t \ \text{and}\ \underline{Y}(s),\ s < t]$$

$$(19)\qquad = A_{YY}(1)\ \underline{Y}(t-1) + \ldots + A_{YY}(m)\ \underline{Y}(t-m) + A_{YX}(1)\ \underline{X}(t-1) + \ldots + A_{YX}(m)\ \underline{X}(t-m) + \eta_Y^*(t) \quad .$$

One often seeks a parsimonious parameterization of the filter with output $\underline{Y}^*(\cdot)$. It might be sought through stepwise regression among predictor formulas of the form

$$(20)\qquad \underline{Y}^*(t) = C_{YY}(1)\ \underline{Y}(t-1) + \ldots + C_{YY}(m)\ \underline{Y}(t-m) + C_{YX}(1)\ \underline{X}(t-1) + \ldots + C_{YX}(m)\ \underline{X}(t-m) + C_{Y\eta_Y}(1)\ \eta_Y(t-1) + \ldots + C_{Y\eta_Y}(m)\ \eta_Y(t-m) \quad .$$

Once a relation of form (20) has been fitted, it can be computed recursively.

The foregoing models correspond to time domain versions of predictor formulas for $\underline{Y}(t)$ in which no rank constraints are imposed on the matrix coefficients. It would be of interest to develop time domain versions of the results of Brillinger (1969) on predictors with rank restraints.

(5) A final point, which seems to me the most important: multivariate analysis of the joint innovation covariance matrix

$$(21)\qquad \tilde{\Sigma} = E[\underline{\underline{\eta}}(t)\ \underline{\underline{\eta}}'(t)]$$

provides interesting relations among components of the time series. The use of regression analysis of $\tilde{\Sigma}$ was discussed in equations (17) - (19). The eigenvalues and eigenvectors of $\tilde{\Sigma}$ seem worth being routinely

computed and examined to indicate ways of reducing the dimensionality of the data vector. The canonical correlations between η_X and η_Y seem to have meaningful interpretations, such as for testing for lack of correlation between $\underline{X}(\cdot)$ and $\underline{Y}(\cdot)$.

5. Autoregressive Approach to a Single Time Series

When modelling a single time series $X(\cdot)$, one is interested in describing the innovation to series filter [which transforms the innovation $\eta(\cdot)$ to $X(\cdot)$] in time domain terms.

Any filter can be approximately expressed as a combination of autoregressive and moving average terms:

$$
\begin{aligned}
&X(t) + a_1 X(t-1) + \ldots + a_m X(t-m) \\
(1) \qquad &= \eta(t) + b_1 \eta(t-1) + \ldots + b_q \eta(t-q) \quad ,
\end{aligned}
$$

one regards the orders m and q, as well as the coefficients $a_1, \ldots, a_m, b_1, \ldots, b_q$ as parameters to be estimated. In this formula, it is usual to think of $\eta(\cdot)$ as a white noise series. A minor point of this paper is: we always require $\eta(\cdot)$ to be the innovation series of $X(\cdot)$.

It turns out that assuming the model for $X(\cdot)$ to be of the form (1) is wquivalent to assuming a model for the one step linear predictor $X^*(t)$ of the form

$$
\begin{aligned}
&X^*(t) = a_1 X(t-1) + \ldots + a_m X(t-m) \\
(2) \qquad &\qquad + b_1 \eta(t-1) + \ldots + b_q \eta(t-q) \quad .
\end{aligned}
$$

When $X^*(t)$ satisfies model (2) with: (i) $b_1 = \ldots = b_q = 0$, we call $X(\cdot)$ an autoregressive scheme of order m, (ii) $a_1 = \ldots = a_m = 0$, we call $X(\cdot)$ a moving average scheme of order q, (iii) some a's and some b's non-zero, we call $X(\cdot)$ a mixed scheme.

In other words, the usual models considered for stationary time

series can equivalently be formulated as models for the predictors $X^*(t)$.

In modelling time series our aim is to obtain a parsimonious parameterization of the form (2). There are available several methods for estimating the parameters of the mixed scheme (2); see Box and Jenkins (1969), Durbin (1959), (1960), Walker (1961), (1962), (1967), and Philips (1967). Possibly a new variant is the following method.

First, fit $X(\cdot)$ by a long autoregressive scheme

(3) $$X(t) = c_1 X(t-1) + \dots + c_M X(t-M) + \eta(t) .$$

Efficient estimators $\hat{c}_1, \dots, \hat{c}_M$ of the coefficients of autoregressive scheme can be computed (at little computational costs by recursive methods).

Second, consider the transfer function

(4) $$C(z) = 1 - c_1 z - c_2 z^2 - \dots - c_M z^M$$

and form its estimator

(5) $$\hat{C}(z) = 1 - \hat{c}_1 z - \hat{c}_2 z^2 - \dots - \hat{c}_M z^M .$$

Third, note that the transfer functions which we seek to estimate

(6) $$A(z) = 1 - a_1 z - \dots - a_m z^m , \quad B(z) = 1 + b_1 z + \dots + b_q z^q$$

are related to $C(z)$ by

(7) $$\frac{1}{C(z)} = \frac{B(z)}{A(z)} \quad \text{or} \quad C(z) = \frac{A(z)}{B(z)} .$$

The parameters of $A(z)$ and $B(z)$ are to be estimated (by non linear least squares) from

(8) $$\hat{C}(e^{i\omega}) = \frac{A(e^{i\omega})}{B(e^{i\omega})} + \text{error}(e^{i\omega})$$

where the error term is a time series (regarded as a function of the index ω) defined by

(9) $$\text{error}(e^{i\omega}) = \hat{C}(e^{i\omega}) - C(e^{i\omega}) \quad .$$

Point Ten of this paper is: it can be shown that the error series (9) can be regarded as asymptotically uncorrelated at different frequencies and with variance

(10) $$\text{Var}[\hat{C}(e^{i\omega})] = \frac{M}{T} |C(e^{i\omega})|^2$$

which is easily estimated.

To motivate (10) let us note that one may regard the estimated autoregressive coefficients

(11) $$\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_M$$

as a "covariance stationary time series" with means $c_1, \ldots, c_M$ and spectral density function

(12) $$f_C(\omega) = \frac{1}{2\pi} \frac{1}{T} |C(e^{i\omega})|^2 \quad .$$

By the theory of the periodogram

(13) $$\begin{aligned} E|\hat{C}(e^{i\omega}) - C(e^{i\omega})|^2 &= E \left| \sum_{k=1}^{M} (\hat{c}_k - c_k) e^{i\omega k} \right|^2 = 2\pi M f_C(\omega) \\ &= \frac{M}{T} |C(e^{i\omega})|^2 \end{aligned}$$

and the values $\hat{C}(e^{i\omega})$ at different frequencies are asymptotically

uncorrelated.

Point Eleven of this paper is: fitting a suitably long autoregressive scheme to a univariate stationary time series is a possible method of spectral density estimation [especially when one assumes that there are no lines in the spectrum].

The usual type of estimator of the normalized spectral density $\bar{f}(\omega)$ [where $\bar{f}(\omega) = f(\omega)/R(0)$] of a stationary time series is a filtered sample spectral density of the form

$$\bar{f}_{T,M}(\omega) = \frac{1}{2\pi} \sum_{|v|<T} \cos v\omega \, k(\tfrac{v}{M}) \, \rho_T(v) \tag{14}$$

where M is a suitable integer (called the truncation point) and $k(\cdot)$ is a suitable kernel. There is an extensive literature on how to choose M and $k(u)$ [see Parzen (1967b) and (1967c)].

It appears that an alternative estimator which is bias free and has similar variability is the autoregressive spectral estimator, defined by

$$\hat{\bar{f}}_M(\omega) = \frac{1}{2\pi} \hat{\sigma}^2_M \left|1 - \sum_{k=1}^{M} \hat{c}_k \, e^{-i\omega k}\right|^{-2} . \tag{15}$$

While the idea of estimating the spectral density by first fitting an autoregressive scheme has been alluded to in the literature, there has been no treatment of its asymptotic variance. A variability theory is now being developed by Parzen (1969). The properties of (15) are discussed at the end of the next section in the context of the multivariate case.

Finally we briefly discuss the question of how to determine the order M of a suitably long autoregressive scheme to fit to a sample

$\{X(t),\ t=1,\ 2,\ \ldots,\ T\}$ of a zero mean covariance stationary time series.

For each order m, one can recursively form: (i) estimators $\hat{a}_{1,m}, \ldots, \hat{a}_{m,m}$ of the coefficients of the predictor of finite memory m

$$(16)\quad E[X(t)|X(t-1), \ldots, X(t-m)] = a_{1,m} X(t-1) + \ldots + a_{m,m} X(t-m) ,$$

and (ii) the sample innovation variance of order m

$$(17)\qquad \hat{\sigma}^2_m = \{\rho_T(0) - \hat{a}_{1,m}\,\rho_T(1) - \ldots - \hat{a}_{m,m}\,\rho_T(m)\}\ R_T(0)$$

where $\rho_T(v)$ is the sample correlation function. Define

$$(18)\qquad \lambda_m = -T \log \hat{\sigma}^2_m \quad ;$$

it is increasing as a function of $m = 1, 2, \ldots, T-1$ and asymptotes to $\lambda_\infty = -T \log \tilde{\sigma}^2_\infty$, where $\tilde{\sigma}^2_\infty$ is the infinite memory prediction error variance.

A procedure for choosing an appropriate order M such that $X(\cdot)$ can be regarded as an autoregressive scheme of order M is: Choose M to be the smallest value of m such that $\lambda_\infty - \lambda_m$ is less than the 95% significance value of the Chi-square distribution with $T-m$ degrees of freedom. Extensive investigation is needed on the theory and application of this procedure.

This suggestion can be roughly justified by the theory of likelihood ratio tests of the hypothesis that the series satisfies an autoregressive scheme of order m versus the alternative hypothesis that it satisfies an autoregressive scheme of order $T-1$ [see Whittle (1952) or Whittle's appendix to Wold (1954)].

It seems to me also justified from the likelihood point of view

since the likelihood of the data under the parameters $\hat{a}_{1,M}, \ldots, \hat{a}_{M,M}$ can be considered to be not "significantly" different from the maximum likelihood of the data (which is a monotone function of λ_∞).

6. Multiple Spectral Density Estimation

In this section we discuss autoregressive approaches to estimating the spectral density matrix

$$(1) \qquad \underline{f}(\omega) = \begin{bmatrix} f_{11}(\omega) & \cdots & f_{1r}(\omega) \\ \cdots & \cdots & \cdots \\ f_{r1}(\omega) & \cdots & f_{rr}(\omega) \end{bmatrix} .$$

Traditionally one estimates $\underline{f}(\omega)$ by estimating each entry $f_{hj}(\omega)$ by a filtered sample cross-spectral density

$$(2) \qquad \hat{f}_{hj;T,M}(\omega) = \frac{1}{2\pi} \sum_{|v|<T} e^{-iv\omega} k(\frac{v}{M}) R_{hj;T}(v) ,$$

where $R_{hj;T}$ is the sample cross-covariance function. Except for ease of developing the distribution theory of the estimator $\hat{\underline{f}}(\omega)$, there seems to be no reason why one should use the <u>same</u> truncation point for each component $f_{hj}(\omega)$.

A method of letting the data determine an appropriate truncation point for each component is to estimate it via a sample analogue of the formula

$$(3) \qquad \begin{aligned} f_{hj}(\omega) &= \frac{1}{2} \{f_{X_h+X_j}(\omega) - f_{X_h}(\omega) - f_{X_j}(\omega)\} \\ &\quad + \frac{i}{2} \{f_{X_h+iX_j}(\omega) - f_{X_h}(\omega) - f_{X_j}(\omega)\} . \end{aligned}$$

Each univariate spectral density which appears in this formula could be estimated by autoregressive spectral estimation, although further research is needed on the theory of the complex valued univariate series

$X_h(\cdot) + iX_j(\cdot)$.

In the literature of empirical spectral analysis, remarks are frequently made about the value of prewhitening or prefiltering. It has long been my view that <u>prewhitening is of value but only when guided by model building</u>. An important approach to estimation of the spectral density matrix which could be said to use prewhitening is as follows.

First, generate the <u>individual</u> innovations $\eta_j(\cdot)$ of each time series $X_j(\cdot)$ by fitting to it a suitably long autoregressive scheme:

$$\eta_j(t) = X_j(t) - c_1^{(j)} X_j(t-1) - \ldots - c_{M_j}^{(j)} X_j(t-M_j) \quad . \tag{4}$$

Second, by some method of spectral density matrix estimation form estimators of the spectral density matrix $\{f_{\eta_h \eta_j}(\omega)\}$ of the multiple time series of individual innovations.

Third, estimate $f_{X_h X_j}(\omega)$ by

$$\begin{aligned} \hat{f}_{X_h X_j}(\omega) = \hat{f}_{\eta_h \eta_j}(\omega) \, &\{1 - c_1^{(h)} e^{-i\omega} - \ldots - c_{M_h}^{(h)} e^{-i\omega M_h}\} \\ &\{1 - c_1^{(j)} e^{i\omega} - \ldots - c_{M_j}^{(j)} e^{i\omega M_j}\} \quad . \end{aligned} \tag{5}$$

A method of spectral density matrix estimation whose value remains to be investigated is via the <u>joint autoregressive estimator</u> which is formed by first fitting a vector autoregressive scheme to the multiple time series $\underline{X}(\cdot)$. To conclude this paper we describe the main features of this approach.

In the notation introduced in Section 2, the joint autoregressive cross-spectral density matrix estimator is defined by

(6) $$\hat{\underline{f}}_M(\omega) = \frac{1}{2\pi} \{\hat{A}(-\omega)\, \hat{\tilde{\Sigma}}^{-1}\, \hat{A}(\omega)'\}^{-1}$$

where

(7) $$\hat{A}(\omega) = I - \hat{A}(1)e^{-i\omega} - \dots - \hat{A}(M)e^{-i\omega M} .$$

The order M used would be determined by a goodness of fit test for joint autoregressive schemes fitted to a multiple time series [see Whittle (1953)]. Here we are interested only in the variability of the estimator $\hat{f}_M(\omega)$ under the assumption that $\underline{X}(\cdot)$ is a zero mean stationary multiple time series satisfying a joint autoregressive scheme of order M.

Research is in progress to prove (as rigorously as possible) that for $0 < \omega < \pi$, with

(8) $$\nu = \frac{T}{M}\,\frac{1}{2} ,$$

$\nu\, \hat{\underline{f}}_M(\omega)$ has a complex Wishart distribution of dimension r, degrees of freedom ν, and covariance matrix $\underline{f}(\omega)$.

The interpretation of this result is that the variability properties of the autoregressive cross-spectral estimator $\hat{\underline{f}}_M(\omega)$ are the same as those of the filtered sample spectral density matrix with rectangular kernel

(9) $$\begin{aligned} k(u) &= 1 , \quad 0 \le |u| \le 1 \\ &= 0 , \quad |u| > 1 \end{aligned}$$

for which $\int_{-\infty}^{\infty} k^2(u)du = 2$.

Some advantages of the autoregressive cross-spectral estimator

seem to me to be:

(1) No window is involved in forming $\hat{\underline{f}}_M(\omega)$, so we avoid a debate as to the choice of window.

(2) The truncation point M can be chosen on the basis that the multiple time series $\underline{X}(\cdot)$ passes a goodness of fit test for obeying a joint autoregressive scheme of order M.

(3) Under the assumption that the multiple time series $\underline{X}(\cdot)$ obeys a joint autoregressive scheme of order M, the autoregressive cross-spectral estimator has much smaller bias than filtered sample cross-spectral density estimators.

(4) Autoregressive cross-spectral estimators are easily <u>updated</u> for additional observations and therefore lend themselves to adaptive estimation [compare Jones (1966)].

REFERENCES

Akaike, H. and Yamanouchi, Y. (1963). On the statistical estimation of frequency response function. Ann. Inst. Statist. Math. 14, 23-56.

Anderson, T. W. (1958). Introduction to Multivariate Statistical Analysis, Wiley: New York.

Box, G. M. and Jenkins, G. M. (1969). Statistical Models for Forecasting and Control, Holden Day: San Francisco.

Brillinger, D. (1969). The canonical analysis of stationary time series. In Second International Symposium on Multivariate Analysis, edited by P. R. Krishnaiah, Academic Press: New York.

Brillinger, D. R. (1970). An Introduction to the Frequency Analysis of Vector-valued Time Series, Holt, Rinehart and Winston: New York, to be published.

Durbin, J. (1959). Efficient estimation of parameters in moving-average models. Biometrika 46, 306-316.

Durbin, J. (1960). The fitting of time-series models. Review International Statistical Institute 28, 233-244.

Gentleman, W. M. and Sande, G. (1966). Fast Fourier transforms - for fun and profit. Proceedings Fall Joint Computer Conference, 1966, 563-578.

Goodman, N. R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). Ann. Math. Statist. 34, 152-177.

Granger, C. W. J. (1964). Spectral Analysis of Economic Time Series, Princeton University Press: Princeton.

Hamon, B. V. and Hannan, E. J. (1963). Estimating relations between time series. J. of Geophysical Research 68, 6033-6041.

Hannan, E. J. (1963). Regression for time series. In Time Series Analysis, edited by M. Rosenblatt, Wiley: New York.

Hannan, E. J. (1965). The estimation of relationships involving distributed lags. Econometrika 33, 206-224

Hannan, E. J. (1967). The estimation of a lagged regression relation. Biometrika 54, 409-418.

Jenkins, G. M. (1965). A survey of spectral analysis. Applied Statist. 14, 2-32.

Jenkins, G. M. and Watts, D. (1968). Spectrum Analysis and Its Applications, Holden Day: San Francisco.

Jones, R. H. (1964). Prediction of multivariate time series. J. of Applied Meteorology 3, 285-289.

Jones, R. H. (1966). Exponential smoothing for multivariate time series. J. Roy. Statist. Soc. Ser. B 28, 241-251.

Kendall, M. G. and Stuart, A. (1966). The Advanced Theory of Statistics, Vol. 3, Griffin: London.

Koopmans, L. (1964). On the multivariate analysis of weakly stationary stochastic processes. Ann. Math. Statist. 35, 1765-1780.

Mann, H. B. and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. Econometrika 11, 173-220.

Masani, P. (1966). Recent trends in multivariate prediction theory, pp. 351-382 of Multivariate Analysis, edited by P. R. Krishnaiah, Academic Press: New York.

Nettheim, N. (1966). The Estimation of Coherence, Ph.D. Thesis, Statistics Department, Stanford University.

Parzen, E. (1962). Stochastic Processes, Holden Day: San Francisco.

Parzen, E. (1967a). The role of spectral analysis in time series analysis. Rev. Inst. Internat. Statist. 35, 125-141.

Parzen, E. (1967b). On empirical multiple time series analysis, pp. 305-340 in Proc. Fifth Berkeley Symp., Vol. 1, edited by L. LeCam, University of California Press: Berkeley.

Parzen, E. (1967c). Time Series Analysis Papers, Holden Day: San Francisco.

Parzen, E. (1968). Statistical Spectral Analysis (Single-channel Case) in 1968, Stanford University Statistics Department, Technical Report No. 11.

Parzen, E. (1969). Empirical Time Series Analysis, Holden Day: San Francisco.

Philips, A. W. (1966). Estimation of Systems of Difference Equations with Moving Average Disturbances, invited address at Econometric Society meeting in San Francisco, 27 December 1966.

Priestley, M. (1965). Evolutionary spectra and non-stationary processes. J. Roy. Statist. Soc, Ser. B. 27, 204-237.

Robinson, E. A. (1967). Multichannel Time Series Analysis with Digital Computer Programs, Holden Day: San Francisco.

Rosenblatt, M. (1959). Statistical analysis of stochastic processes with stationary residuals, Probability and Statistics (Cramér Volume), edited by U. Grenander, Wiley: New York, 300-330.

Rozanov, Yu. A. (1967). Stationary Stochastic Processes, Holden Day: San Francisco.

Tick, L. J. (1967). Estimation of coherency, Advanced Seminar on Spectral Analysis of Time Series, edited by B. Harris, Wiley: New York.

Wahba, G. (1968). On the distribution of some statistics useful in the analysis of jointly stationary time series. Ann. Math. Statist., to appear.

Walker, A. M. (1961). Large sample estimation of parameters for moving-average models. Biometrika 48, 343-357.

Walker, A. M. (1962). Large-sample estimation of parameters for auto-regressive processes with moving-average residuals. Biometrika 49, 117-131.

Walker, A. M. (1967). Some tests of separate families of hypotheses in time series analysis. Biometrika 54, 39-68.

Whittle, P. (1952). Tests of fit in time series. Biometrika 39, 309-318.

Whittle, P. (1953). The analysis of multiple stationary time series. J. Roy. Statist. Soc. Ser. B 15, 125-139.

Whittle, P. (1963a). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. Biometrika 50, 129-134.

Whittle, P. (1963b). Prediction and Regulation, English Universities Press: London.

Wold, H. (1954). A Study in the Analysis of Stationary Time Series, Second Edition, Uppsala.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Stanford University Department of Statistics Stanford, California	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

MULTIPLE TIME SERIES MODELLING

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report, July, 1968

5. AUTHOR(S) (Last name, first name, initial)

Parzen, Emanuel

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
July 8, 1968	34	44

8a. CONTRACT OR GRANT NO. DA-ARO(D)-31-124-G726 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Report No. 22 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) Tech. Report No. 12 on Nonr 225(80).

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Army Research Office Durham, North Carolina

13. ABSTRACT

This paper seeks to provide a general framework for the theory and practice of multivariate analysis of time series. It seeks to compare:

(1) Spectral approaches to finding relations among time series.

(2) Time domain or innovations approaches to finding relations among time series.

The paper also seeks to focus attention on:

(3) Innovations approaches to cross-spectral estimation.

(4) The problem of multivariate analysis of the joint innovations covariance matrix and the sampling properties of its estimators.

The various sections are entitled: 2. Innovation Approaches to Modelling, 3. Spectral Approaches to Modelling, 4. Relations Between Time Series, 5. Autoregressive Approach to a Single Series, 6. Multiple Spectral Density Estimation.

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

Unclassified

Security Classification

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Stationary Time Series						
Spectral Analysis						
Multivariate Analysis						
Autoregressive Time Series						
System Identification						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
_______________________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
_______________________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
_______________________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

RECENT REPORTS ON TIME SERIES ANALYSIS

Research in time series analysis under the direction of Professor Emanuel Parzen is supported by the Office of Naval Research - Contract Nonr 225(80), and the Army Research Office - Grant ARO(D)-31-124 G726.

Since July 1, 1966, the following technical reports have been issued:

Author	Title	Date	Agency	Report No.
J. W. Van Ness	Empirical Nonlinear Prediction and Polyspectra	9-12-66	ARO NSF	18 16
E. Sobel	Prediction of a Noise-distorted, Multivariate, Non-stationary Signal	10-31-66	ARO	19
G. Wahba	On the Joint Distribution of Circulant Forms in Jointly Stationary Time Series	1-16-67	NSF	17
E. Parzen	Time Series Analysis for Models of Signal Plus White Noise	9-12-66	ONR	6
E. Parzen	Analysis and Synthesis of Forecasting Models for Time Series	1-16-67	ONR	7
G. Wahba	Some Tests of Independence for Stationary Multivariate Time Series	9-15-67	ONR	8
H. Akaike	Low Pass Filter Design	11-30-67	ARO ONR	20 9
M. Brown	Convergence in Distribution of Stochastic Integrals	1-30-68	ONR	10
R. Olshen	Representing Finitely Additive Invariant Probabilities	3-15-68	ARO	21
E. Parzen	Statistical Spectral Analysis (Single Channel Case) in 1968	6-10-68	ONR	11
E. Parzen	Multiple Time Series Modelling	7-8-68	ONR ARO	12 22
A. M. Walker	Large Sample Properties of Harmonic Components in a Time Series with Stationary Residuals. I. Independent Residuals	7-8-68	ARO	23
P. Shaman	Cross-spectral Relationships in a Linear Time Series Model	In Preparation	ARO ONR	24 13
P. Shaman	On the Inverses of the Covariance Matrix of a First Order Moving Average	In Preparation	ARO	25